13th International Conference on Applied Informatics
Eszterházy Károly Catholic University
Eger, Hungary, February 19–21, 2026

Application of Large Language Models on Structured Corporate Data

István Szeglet, Norbert Pataki

Department of Programming Languages and Compilers
Eötvös Loránd University
szeglet.istvan@gmail.com, patakino@elte.hu

Abstract

The Large Language Models (LLMs) are rapidly evolving tools, they went through significant progression. Nowadays LLMs are utilized for text comprehension, summarization, or translation and chatbots. These models work outstandingly in environments of natural languages. However, their work is not free of bugs and mistakes. Moreover, the corporate data is not available in free text, but in a structured format, for instance, relational databases, financial statements, balance sheets, Enterprise Resource Planning (ERP) systems, business reports. The efficient involvement of LLMs is still challenging, especially when the accuracy and traceability are critical requirement [4].

In this paper, we present methods for connecting LLMs and structured corporate data in a way that enhances the decision-making with added value in a constructive way. The focus is not only on the query of the data, but on its interpretation and explanation. In this paper, we discuss three main approaches: the schema-oriented prompting and related prompting strategies, the automatic generation of structured queries, and the Retrieval-Augment Generation-based (RAG-based) architectures.

One of the highest risk in the corporate environment is the appearance of hallucinations when the model produces factually incorrect information with confident tone [2]. This phenomenon is rather critical in financial, accounting data, or other area of application with high risk where even only one inexact statement may result in significant consequences. In this research, we deal with special focus to the techniques that reduce the appearance of hallucinations. For instance, we take advantage of explicit resource usage, deterministic query steps and guarantees for

the validation.

The essence of the schema-oriented prompting is that model takes the question in an unblind-way, but it receives information about the structure of database beforehand. With the knowledge of tables, fields and connections, the LLM is able to interpret the user's intention that results in reduced chance to generate erroneous queries or misleading answers. The results present that rather simple schema descriptions improve the quality of SQL queries significantly.

The transformation of natural language queries to a structured one plays a key role in the utilization of LLMs in corporate environment. The paper presents that the model is the most efficient when not the final answer is prompted directly, but at first a formal executable query is produced. This approach increases the transparency and makes the intermediate steps validable.

The Retrieval-Augmented Generation-based system provide further advances by finding relevant data from external source before generation [3]. These data can be structured tables, reports, documentations, so the model relies on not only the learned patterns, but current context-aware corporate data as well [1]. This approach results in more precise answers, moreover, it can highlight the sources of answers.

The data security and authorization play an important role during the design of LLM-based systems that work with structured corporate data. An enterprise environment often utilizes authorization levels and sensitive financial information and its special requirements are in-use that restrict the direct utilization of data.

A further important aspect is the interpretability of the LLM-based solutions. In the corporate decision-making, generation of a simple answer or result is not enough, but its justification and contextualization are required as well. The presented approaches support that the data and intermediate steps behind the generated answers be transparent. These features increase the users'confidence and the practical acceptability of the systems.

We evaluate the approaches through corporate use-cases, such as automatic report generation that LLM generates comprehensible summaries and reports, or explanation of anomalies that modell assists in the interpretation of the reasons behind weird differences. Hybrid LLM-database solutions can improve the data accessibility and its business interpretability significantly.

As a summary, the large language models only cannot replace the traditional enterprise business solutions and software systems. However, proper integration of the LLM can become valuable auxiliary solution. In the future, the most important role of corporate LLMs is not the generation of data, but its explanation and connections.

References

[1] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, M. Wang, H. Wang: *Retrieval-Augmented Generation for Large Language Models: A Survey*, 2024, doi: 10.48550/arXiv.2312.10997, arXiv: 2312.10997 [cs.CL], url: https://arxiv.org/abs/2312.10997.

[2] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, P. Fung: *Survey of Hallucination in Natural Language Generation*, ACM Comput. Surv. 55.12 (Mar. 2023), issn: 0360-0300, doi: 10.1145/3571730, url: https://doi.org/10.1145/3571730.

[3] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, D. Kiela: *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*, 2021, doi: 10.48550/arXiv.2005.11401, arXiv: 2005.11401 [cs.CL], url: https://arxiv.org/abs/2005.11401.

[4] Z. Yang, Q. Xu, S. Gao, C. Yang, G. Wang, Y. Zhao, F. Kong, H. Liu, W. Wang, J. Xiao: *OceanBase Paetica: A Hybrid Shared-Nothing/Shared-Everything Database for Supporting Single Machine and Distributed Cluster*, Proc. VLDB Endow. 16.12 (Aug. 2023), pp. 3728–3740, issn: 2150-8097, doi: 10.14778/3611540.3611560, url: https://doi.org/10.14778/3611540.3611560.